Revised copy 1/67

APPLICATION OF SPECTRAL THEORY

FOR FINDING FUNCTIONS OF SQUARE MATRICES

BU-230-M

Abdossamad Hedayat
Biometrics Unit
Cornell University
Ithaca, New York

December, 1966

ABSTRACT

The concept of a defined function on the spectrum of a given square matrix A is given. To find any defined function such as $f(A) = e^A$ or A^K, K positive integer, all we need is an annihilating polynomial for A. The procedure for finding the general form of f(A) is also given. This procedure follows from the concept that an $n \times n$ matrix can be regarded as a vector in a space of V^{n^2} of dimension n^2. If minimal polynomial for the given A is of degree m then all the defined functions of A form an m dimensional subspace of V^{n^2}. Methods for finding a basis for this subspace and coordinate vector for f(A) are given. These concepts have application for finding the power of transition probability matrices and solving the linear systems of differential equations. Finally several useful theorems and corollaries related to exponential and trigonometric functions of a square matrix A are stated.

APPLICATION OF SPECTRAL THEORY
FOR FINDING FUNCTIONS OF SQUARE MATRICES

BU-230-M Abdosammad Hedayat December, 1966

INTRODUCTION:

Since we are going to give the definition of a defined function of square matrix A through the behavior of this function on the point of the spectrum of A, and because our procedures will be based on the spectral theory, hence we prefer at first to give the definitions and theorems related to this subject. Two methods for finding the functions of square matrices will be given. The first method is useful if our primary interest is to find a single function of the given square matrix A. The second method is more general and will be helpful if we want to find several functions of the given square matrix A. The second procedure will be based on constructing a basis for the subspace of all defined functions of square matrix A and related coordinate vectors.

DEFINITION:

A scalar polynomial $f(\lambda)$ is called an annihilating polynomial of the square matrix A if $f(A) = 0$. $f(\lambda)$ is not unique.

Ex. $f(\lambda) = \lambda^4 - 3\lambda^3 - 9\lambda^2 - 5\lambda$ is such a polynomial for $\begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$

DEFINITION:

An annihilating polynomial $m(\lambda)$ of least degree with highest coefficient 1 is called the minimal polynomial of A. $m(\lambda)$ is unique and divides any annihilating polynomial for A without remainder.

Ex. $m(\lambda) = \lambda^2 - 1$ is the minimal polynomial for $A = \begin{bmatrix} -1 & -2 & -2 \\ 1 & 2 & 1 \\ -1 & -1 & 0 \end{bmatrix}$

DEFINITION

The scalar polynomial $c(\lambda) = |A-\lambda I|$ is called the characteristic polynomial of the matrix A. The set of distinct roots of $c(\lambda)$ is called spectrum of A.

Ex. $c(\lambda) = (\lambda+1)(\lambda-1)^2$ is the characteristic polynomial for the matrix A in the second definition.

Cayley-Hamilton Theorem:

If

$$c(\lambda) = |A-\lambda I| = \sum_{i=0}^{n} (-1)^{n-i} a_i \lambda^{n-i}$$

$$\Rightarrow \sum_{i=0}^{n} (-1)^{n-i} a_i A^{n-i} = 0 \quad .$$

Corollary 1 to the C. H. Theorem:

Any positive integral power A^K of the square matrix A of order n for $K \geq n$ is linearly expressible in terms of the unit matrix of order n and the first (n-1) powers of A. i.e.

$$A^K = a_0 I + a_1 A + \cdots + a_{n-1} A^{n-1} \quad , \quad k \geq n \ .$$

Ex. $$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} , \ A^K = 0 \cdot I + (2^{2-K}-1)A + (2-2^{2-K})A^2$$

$$= \begin{bmatrix} 2^{-2}+2^{-K+1} & \frac{1}{2} & 2^{-2}-2^{-K-1} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 2^{-2}-2^{-K-1} & \frac{1}{2} & 2^{-2}+2^{-K-1} \end{bmatrix} .$$

Corollary 2 to the C. H. Theorem:

Any negative integral power A^{-K} of the non-singular square matrix of order n is linearly expressible in terms of the unit matrix and the first (n-1) powers of A. i.e.

$$A^{-K} = b_0 I + b_1 A + \cdots + b_{n-1} A^{n-1} \quad , \quad |A| \text{ exist} ,$$

$$A = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix} , \quad A^{-K} = \frac{5.2^{-K}-2.5^{-K}}{3} I + \frac{5^{-K}-2^{-K}}{3} A \ .$$

DEFINITION:

We say a square matrix A has a <u>K-fold degeneracy for the characterieristic value λ_i</u>, if there exist K linearly independent characteristic vectors for λ_i. K is also called the geometric multiplicity of λ_i.

Ex. Matrix A in the second definition of page 1 has 2-fold degeneracy for $\lambda = 1$.

DEFINITION:

If a square matrix $A_{n \times n}$ has a total of n linearly independent characteristic vectors, regardless of degeneracy, then A is said to be semi-simple.

Ex. $A = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 3 & -2 \\ 4 & 1 & 0 \end{bmatrix}$ is not semi-simple .

$A = \begin{bmatrix} 2 & 7 \\ 0 & 3 \end{bmatrix}$ is semi-simple .

DEFINITION:

Generalized characteristic vector: If a square matrix A is not semi-simple, i.e. for λ_i with multiplicity $m_i > 1$, related degeneracy of characteristic vectors associated with λ_i is less than m_i, then those vectors that are annihilated by $(A-\lambda_i I)^2$, $(A-\lambda_i I)^3$, etc., until we get enough vectors, are called generalized characteristic vectors associated with λ_i.

THEOREM:

Every arbitrary square matrix A is always similar to some matrix of the form $A = TJT^{-1}$, where J is called the jordan normal form. Columns of T are constructed from characteristic and generalized characteristic vectors of A. (See Nering, "Linear Algebra and Matrix Theory", for a short and beautiful presentation.)

THEOREM:

A N.S. condition under which a square matrix A is similar to a diagonal matrix $D^{\dagger}$, i.e. A can be expressed in the form of $A = PDP^{-1}$, where D is a diagonal matrix of all characteristic roots: A is semi-simple or multiplicity of all roots in $m(\lambda)$ is 1 or rank of $(A-\lambda_i I) = n - m_i$ for all "i", where m_i is the (Algebraic) multiplicity of λ_i in $c(\lambda)$. In short, Algebraic multiplicity of λ_i = Geometric multiplicity of λ_i.

Ex.

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$ cannot be expressed as $A = PDP^{-1}$ otherwise, $PDP^{-1} = PIP^{-1} = I \neq A$

$$A = \begin{bmatrix} 0 & 1 & 0 & 4 \\ 1 & 0 & -4 & 0 \\ 0 & -4 & 0 & 1 \\ 4 & 0 & 1 & 0 \end{bmatrix}$$ cannot be expressed as $A = PDP^{-1}$

otherwise, $PDP^{-1} = POP^{-1} = 0 \neq A$

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 4 & 0 & -1 \\ 4 & -2 & 1 \end{bmatrix}$$ can be expressed as $A = PDP^{-1}$

As an example

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 \\ -2 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}$$

DEFINITION:

Given $m(\lambda)$ of a matrix A,

$$m(\lambda) = (\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_s)^{m_s} \quad . \qquad (1)$$

---

† If this is the case A is said to be diagonalizable.

We say that $f(\cdot)$ is defined on the spectrum of A if

$$f(\lambda_i), f'(\lambda_i), \cdots, f^{(m_i-1)}(\lambda_i) \qquad i=1,2,\cdots,s \tag{2}$$

exist (i.e., have meaning).

Ex. 1:

$$\text{For } A = \begin{bmatrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \qquad \text{we have } m(\lambda) = (\lambda-3)^2(\lambda-1)$$

then, $f(x) = e^x$ is defined on the spectrum of A, since

$$f(3) = e^3,\ f'(3) = e^3,\ f(1) = e^1 \text{ are defined.}$$

Ex. 2:

$$\text{For } A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 1 & -1 \end{bmatrix} \qquad \text{we have } m(\lambda) = \lambda^2(\lambda^2+1) \text{ then}$$

$f(x) = \frac{1}{x}$ is not defined on the spectrum of A, since

$f(0) = \frac{1}{0}$ is not defined (has no meaning).

Lemma:

If $g(\cdot)$ and $h(\cdot)$ have the same values on the spectrum of A, then it is easy to see that

$$g(\lambda) = h(\lambda) \ [\text{mod } m(\lambda)] \ . \tag{3}$$

Hence

$$g(A) = h(A) \quad .$$

Ex.

$$A = \begin{bmatrix} 0 & 3 & 4 \\ 0 & 0 & 5 \\ 0 & 0 & 0 \end{bmatrix} \quad , \quad m(\lambda) = \lambda^3$$

Since for $g(x) = x^5$ and $g(y) = y^{10}$ we have $g(0) = h(0) = 0$, $g'(0) = h'(0) = 0$, $g''(0) = h''(0) = 0$. Then

$$g(A) = h(A), \quad \text{i.e.} \quad A^5 = A^{10} \quad .$$

Formal Definition of Function of a Square Matrix

Let $g(\cdot)$ be an arbitrary polynomial that assumes on the spectrum of A the same values as does $f(\cdot)$, then

$$f(A) = g(A) \quad .$$

THEOREM:

Among all the polynomials with complex coefficients that assume on the spectrum of A the same values as $f(\cdot)$ there is one and only one polynomial $g(\cdot)$ that is of degree less than m. This polynomial is uniquely determined by the interpolation conditions:

$$g(\lambda_K) = f(\lambda_K),\ g'(\lambda_K) = f'(\lambda_K),\ \cdots,\ g^{(m_K-1)}(\lambda_K) = f^{(m_K-1)}(\lambda_K)$$

$$(K=1,2,\cdots,s) \quad . \tag{4}$$

The polynomial $g(\cdot)$ is called the Lagrange-Sylvester interpolation polynomial for $f(\cdot)$ on the spectrum of A. We will show later that this is one of the best methods of finding one function of a square matrix at a time.

Corollary:

If A is a square diagonal matrix and f(A) makes sense, then

$$f(A) = \begin{bmatrix} f(\ ) & & 0 \\ 0 & f(\ ) & \\ & & f(\ ) \end{bmatrix}$$

Ex.

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -7 \end{bmatrix}$$

since e^x is defined on the spectrum of A, hence

$$e^A = \begin{bmatrix} e^2 & 0 & 0 \\ 0 & e^5 & 0 \\ 0 & 0 & e^{-7} \end{bmatrix} \quad .$$

Corollary:

If A and B are similar, i.e.

$$A = TBT^{-1} \tag{5}$$

and if f(x) is defined on the spectrum of A, then

$$f(A) = Tf(B)T^{-1} \quad . \tag{6}$$

We see that if A is semi-simple we can select T such that B is diagonal. If B is the jordan normal form for A we still can have a simple expression for $f(B)$.

Computational procedure:

Case 1: Computation of one function at a time.

In this case one of the best ways is using the idea of the theorem on page 6.

Ex. 1:

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 3 & -2 \\ 4 & 1 & 0 \end{bmatrix}$$

What is A^K (K is a positive integer)? We find $g(x)$ such that it agrees with $f(x) = x^K$ on the spectrum of A, since

$$m(\lambda) = c(\lambda) = (\lambda-1)^2(\lambda-3) \quad .$$

$\therefore \quad g(x) = a_0 + a_1 x + a_2 x^2$, hence

$$f(1) = 1^K = g(1) = a_0 + a_1 + a_2$$

$$f'(1) = K \cdot 1^{K-1} = g'(1) = a_1 + 2a_2$$

$$f(3) = 3^K = g(3) = a_0 + 3a_1 + 9a_2$$

$$\Rightarrow a_0 = \frac{3^K-6K+3}{4}, \quad a_1 = \frac{1+4K-3^K}{2}, \quad a_2 = \frac{3^K-2K-1}{4}$$

$\therefore$

$$A^K = \frac{3^K-6K+3}{4} I + \frac{1+4K-3^K}{2} A + \frac{3^K-2K-1}{4} A^2 \quad .$$

Ex. 2:

Find the k^{th} power of the transition probability matrix A.

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$m(\lambda) = \lambda^3 - \frac{3}{2}\lambda^2 + \frac{1}{2}\lambda \ , \quad \lambda_1 = 0, \ \lambda_2 = 0, \ \lambda_3 = \frac{1}{2} \ .$$

With the similar argument as in Example 1

$$f(1) = 1^K = g(1) = a_0 + a_1 + a_2$$

$$f(0) = 0^K = g(0) = a_0 + 0 \cdot a_1 + 0 \cdot a_2$$

$$f(\tfrac{1}{2}) = (\tfrac{1}{2})^K = g(\tfrac{1}{2}) = a_0 + \tfrac{1}{2}a_1 + \tfrac{1}{4}a_2$$

$$\Rightarrow a_0 = 0, \quad a_1 = 2^{2-K}-1, \quad a_2 = 2-2^{2-K} \ .$$

$\therefore$

$$A^K = (2^{2-K}-1)A + (2-2^{2-K})A^2 \ .$$

Ex. 3:

Find the square root of A.

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$$

$$m(\lambda) = c(\lambda) = (2-\lambda)(1-\lambda)$$

$$f(A) = \sqrt{A} \ , \quad \text{let } f(x) = \sqrt{x} \text{ and related } g(x) = a_0 + a_1 x$$

$\therefore$

$$f(1) = \sqrt{1} = g(1) = a_0 + a_1$$

$$f(2) = \sqrt{2} = g(2) = a_0 + 2a_1$$

$\Rightarrow a_0 = 2 - \sqrt{2}\ ,\quad a_1 = \sqrt{2} - 1$

$\therefore$

$$f(A) = \sqrt{A} = (2-\sqrt{2})I + (\sqrt{2}-1)A = \begin{bmatrix} \sqrt{2} & \sqrt{2}-1 \\ 0 & 1 \end{bmatrix}$$

Check:

$$\begin{bmatrix} \sqrt{2} & \sqrt{2}-1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2}-1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} .$$

Ex. 4:

Find cos H and sin H where

$$H = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} . \qquad \text{Note that } H^2 = H.$$

$$m(\lambda) = \lambda(\lambda-1)\ ,\quad \lambda_1 = 0\ ,\quad \lambda_2 = 1\ .$$

a

$f(H) = \cos H$, let $f(x) = \cos x$ and related $g(x) = a_0 + a_1 x$

$\therefore$

$$f(0) = \cos 0 = g(0) = a_0 + 0$$

$$f(1) = \cos 1 = g(1) = a_0 + a_1$$

$$\Rightarrow a_0 = 1\ ,\quad a_1 = (\cos 1) - 1$$

$\therefore$

$$f(H) = \cos H = 1 \cdot I + ((\cos 1)-1)H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + ((\cos 1)-1) \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \cos 1 & 0 & 2((\cos 1)-1 \\ 0 & \cos 1 & 1 - \cos 1 \\ 0 & 0 & 1 \end{bmatrix}$$

b

$f(H) = \sin H$, let $f(x) = \sin x$ and related $g(x) = a_0 + a_1 x$

$\therefore$

$$f(0) = \sin 0 = g(0) = a_0$$

$$f(1) = \sin 1 = g(1) = a_0 + a_1$$

$$\Rightarrow a_0 = 0\ ,\quad a_1 = \sin 1$$

$\therefore$

$$\sin H = \sin 1 \cdot H = \begin{bmatrix} \sin 1 & 0 & 2\sin 1 \\ 0 & \sin 1 & -\sin 1 \\ 0 & 0 & 0 \end{bmatrix} .$$

As a check we like to have $\sin^2 H + \cos^2 H = I$.

$$\sin^2 H = \begin{bmatrix} \sin^2 1 & 0 & 2\sin^2 1 \\ 0 & \sin^2 1 & -\sin^2 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\cos^2 H = \begin{bmatrix} \cos^2 1 & 0 & -2\sin^2 1 \\ 0 & \cos^2 1 & \sin^2 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\Rightarrow \sin^2 H + \cos^2 H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Ex. 5:

Find e^H where $H = \begin{bmatrix} -8 & -18 & -6 \\ 6 & 13 & 4 \\ -6 & -12 & -3 \end{bmatrix}$.

$$m(\lambda) = \lambda(\lambda - 1)\ ,\quad \lambda_1 = 0\ ,\quad \lambda_2 = 1$$

$f(H) = e^H$, let $f(x) = e^x$ and related $g(x) = a_0 + a_1 x$

$\therefore$

$f(0) = e^0 = g(0) = a_0 + 0$

$f(1) = e^1 = g(1) = a_0 + a_1$

$\Rightarrow \quad a_0 = 1 \,, \quad a_1 = e - 1$

$\therefore$

$$e^A = I + (e-1)H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + (e-1) \begin{bmatrix} -8 & -18 & -6 \\ 6 & 13 & 4 \\ -6 & -12 & -3 \end{bmatrix} .$$

Case 2: When it is necessary to deal with several functions of one and the same matrix A, or when the function $f(\lambda)$ depends not only on λ, but also on some parameter t, the following procedure will be convenient.

Since the related notation for the method we are going to give looks complicated, we prefer to give at first the idea through a numerical example. Later in this part the complete formula will be given.

Ex.

Find the general form of the function $f(A)$ for the matrix

$$A = \begin{bmatrix} 3 & -1 & 1 \\ 2 & 0 & 1 \\ 1 & -1 & 2 \end{bmatrix}$$

and hence compute A^K (K positive integer), e^{At}, $\sqrt{A}$ and cos A.

$$m(\lambda) = (\lambda-1)(\lambda-2)^2 \;, \quad \lambda_1 = 1,\ \lambda_2 = 2 \quad \text{with multiplicity 2.}$$

Define $h_{10}(\cdot)$, $h_{20}(\cdot)$, and $h_{21}(\cdot)$ as follows:

$$\left.\begin{array}{ll}
\text{properties of } h_{10}(\cdot) & \left[\begin{array}{l} h_{10}(\lambda_1) = h_{10}(1) = 1 \\ h_{10}(\lambda_2) = h_{10}(2) = 0 \\ h'_{10}(\lambda_2) = h'_{10}(2) = 0 \end{array}\right. \\[2em]
\text{properties of } h_{20}(\cdot) & \left[\begin{array}{l} h_{20}(\lambda_2) = h_{20}(2) = 1 \\ h'_{20}(\lambda_2) = h'_{20}(2) = 0 \\ h_{20}(\lambda_1) = h_{20}(1) = 0 \end{array}\right. \\[2em]
\text{properties of } h_{21}(\cdot) & \left[\begin{array}{l} h'_{21}(\lambda_2) = h'_{21}(2) = 1 \\ h_{21}(\lambda_2) = h_{21}(2) = 0 \\ h_{21}(\lambda_1) = h_{21}(1) = 0 \end{array}\right.
\end{array}\right\} \tag{7}$$

Note the relation between subscript of functions and their values on the spectrum of A .

Now let

$$g(\cdot) = f(\lambda_1)h_{10}(\cdot) + f(\lambda_2)h_{20}(\cdot) + f'(\lambda_2)h_{21}(\cdot) \quad . \tag{8}$$

Note the relation between argument of $f(\cdot)$ and functions $h(\cdot)$. Let us check how $g(\cdot)$ behaves on the spectrum of A.

$$g(\lambda_1) = \underline{g(1)} = f(1)h_{10}(1) + f(2)h_{20}(1) + f'(2)h_{21}(1)$$

$$= f(1)\cdot 1 + f(2)\cdot 0 + f'(2)\cdot 0 = \underline{f(1)}$$

$$g(\lambda_2) = \underline{g(2)} = f(1)h_{10}(2) + f(2)h_{20}(2) + f'(2)h_{21}(2)$$

$$= f(1)\cdot 0 + f(2)\cdot 1 + f'(2)\cdot 0 = \underline{f(2)}$$

$$g'(\lambda_2) = \underline{g'(2)} = f(1)h'_{10}(2) + f(2)h'_{20}(2) + f'(2)h'_{21}(2)$$

$$= f(1)\cdot 0 + f(2)\cdot 0 + f'(2)\cdot 1 = \underline{f'(2)}$$

$\therefore$ by our definition $g(\cdot)$ agrees with $f(\cdot)$ on the spectrum of A, hence

$$f(A) = g(A) = f(\lambda_1)h_{10}(A) + f(\lambda_2)h_{20}(A) + f'(\lambda_2)h_{21}(A)$$

$$= f(1)h_{10}(A) + f(2)h_{20}(A) + f'(2)h_{21}(A) \quad . \tag{9}$$

For the ease of notation let us denote $h_{10}(A)$, $h_{20}(A)$ and $h_{21}(A)$ by Z_{10}, Z_{20}, and Z_{21} respectively. Therefore, by these notations

$$f(A) = f(\lambda_1)Z_{10} + f(\lambda_2)Z_{20} + f'(\lambda_2)Z_{21} \quad . \tag{10}$$

As we see, matrices Z_{10}, Z_{20}, and Z_{21} are independent from the form of $f(\cdot)$ and hence we can compute them once and use them for any function $f(A)$. It can be shown that none of the Z matrices is zero matrix; moreover, they are linearly independent (refer to the last section). For our example

$$f(A) = f(1)Z_{10} + f(2)Z_{20} + f'(2)Z_{21} \quad .$$

Since relation (10) is valid for any defined function $f(\cdot)$ on the spectrum of A and because we have three unknowns Z_{10}, Z_{20}, and Z_{21} we use three simple functions $f_1(\cdot)$, $f_2(\cdot)$, and $f_3(\cdot)$ to construct three equations and hence solve them for Z_{10}, Z_{20}, and Z_{21}. Note that no matter what functions we use as long as it is defined on the spectrum of A we get unique answers for Z matrices. Let

$$f_1(\lambda) = f(\lambda) \equiv 1 \Rightarrow f(A) = I$$

$$f_2(\lambda) = f(\lambda) \equiv \lambda-2 \Rightarrow f(A) = A-2I$$

$$f_3(\lambda) = f(\lambda) \equiv (\lambda-2)^2 \Rightarrow f(A) = (A-2I)^2 \quad .$$

Note also we pick $f_i(\lambda)$, $i=1,2,3$ from the minimal polynomial. You can use any function, but these are usually better choices. Using these functions and substituting them in relation (10) we get

$$I = Z_{10} + Z_{20}$$

$$A-2I = -Z_{10} + Z_{21}$$

$$(A-2I)^2 = Z_{10} \quad .$$

Solving these three equations and three unknowns we find

$$Z_{10} = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix}, \quad Z_{20} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{bmatrix}, \quad Z_{21} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix} .$$

$\therefore$

$$f(A) = f(1)\begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix}, + f(2)\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 1 \end{bmatrix}, + f'(2)\begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix} . \quad (11)$$

Hence, if we define $f(\lambda) = \lambda^K$, $f(\lambda) = e^{\lambda t}$, $f(\lambda) = \lambda^{\frac{1}{2}}$ and $f(\lambda) = \cos\lambda$ we can answer our questions as follows:

$$f(A) = A^K = 1^K \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix} + 2^K \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 1 \end{bmatrix} + K2^{K-1}\begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$f(A) = e^{At} = e^t \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix} + e^{2t}\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 1 \end{bmatrix} + te^{2t}\begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$f(A) = \sqrt{A} = 1^{\frac{1}{2}}\begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix} + 2^{\frac{1}{2}}\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 1 \end{bmatrix} + \frac{1}{2\sqrt{2}}\begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$f(A) = \cos A = \cos(1)\begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix} + \cos(2)\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 1 \end{bmatrix} - \sin(2)\begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

We see how easy one can compute any defined function of A once he gets the general form of f(A).

<u>Remark</u>: Sometimes it is easy to find f(A) once we get relation (10) directly from the following determinant relation.

$$\begin{vmatrix} f(A) & f(1) & f(2) & f'(2) \\ f_1(A) & f_1(1) & f_1(2) & f_1'(2) \\ f_2(A) & f_2(1) & f_2(2) & f_2'(2) \\ f_3(A) & f_3(1) & f_3(2) & f_3'(2) \end{vmatrix} = 0 \quad .$$

Subsitiuting related values for $f_i(\cdot)$ as we found on page 14 we get

$$\begin{vmatrix} f(A) & f(1) & f(2) & f'(2) \\ I & 1 & 1 & 0 \\ A-2I & -1 & 0 & 1 \\ (A-2I)^2 & 1 & 0 & 0 \end{vmatrix} = 0 \quad .$$

Evaluating this determinant we get the same answer for f(A) as we got in relation (11). Later we give enough examples to make the above procedure clear.

<u>Generalization of the Procedure</u>:

Suppose for the given square matrix A, the minimal polynomial is

$$m(\lambda) = (\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2} \cdots (\lambda-\lambda_s)^{m_s} .$$

To find f(A), define g(·) as follows:

$$g(\cdot) = \sum_{i=1}^{s} [f(\lambda_i)h_{i0}(\cdot) + f'(\lambda_i)h_{i1}(\cdot) + \cdots + f^{(m_i-1)}(\lambda_i)h_{im_i-1}(\cdot)] \qquad (12)$$

where functions $h_{ij}(\cdot)$ are defined on the spectrum of A as follows:

$$\begin{cases} h_{i0}(\lambda_i) = 1 \\ h'_{i0}(\lambda_i) = h''_{i0}(\lambda_i) = \cdots = h_{i0}^{(m_i-1)}(\lambda_i) = 0 \\ h_{i0}(\lambda_j) = h'_{i0}(\lambda_j) = \cdots = h_{i0}^{(m_j-1)}(\lambda_j) = 0 \quad \text{for } j=1,2,\cdots,s . \\ i \neq j \end{cases}$$

$$\begin{cases} h'_{i1}(\lambda_i) = 1 \\ h_{i1}(\lambda_i) = h''_{i1}(\lambda_i) = \cdots = h_{i1}^{(m_i-1)}(\lambda_i) = 0 \\ h_{i1}(\lambda_j) = h'_{i1}(\lambda_j) = \cdots = h_{i1}^{(m_j-1)}(\lambda_j) = 0 \quad \text{for } j=1,2,\cdots,s . \\ i \neq j \end{cases}$$

in general $h_{i}n_{i}(\cdot)$ is defined as follows:

$$\begin{cases} h_{in_i}^{(n_i)}(\lambda_i) = 1 \\ h_{in_i}(\lambda_i) = h'_{in_i}(\lambda_i) = \cdots = h_{in_i-1}^{(n_i-1)}(\lambda_i) = h_{in_i+1}^{(n_i+1)}(\lambda_i) = \cdots = h_{in_i}^{(m_i-1)}(\lambda_i)=0 \\ h_{in_i}(\lambda_j) = h'_{in_i}(\lambda_j) = \cdots = h_{in_i}^{(m_j-1)}(\lambda_j) = 0 \quad \text{for } j=1,2,\cdots,s \;.. \\ i \neq j \end{cases} \tag{13}$$

With the same argument as before, we can show that $g(\cdot)$ agrees with $f(\cdot)$ on the spectrum of A. Therefore, $f(A) = g(A)$; therefore,

$$f(A) = g(A) = \sum_{i=1}^{s} [f(\lambda_i)h_{i0}(A) + f'(\lambda_i)h_{i1}(A) + \cdots + f^{(m_i-1)}(\lambda_i)h_{im_i-1}(A)] \;. \tag{14}$$

Now we can select appropriate functions from the minimal polynomial and find matrices $h_{i0}(A),\cdots,h_{im_i-1}(A)$ from the system of linear equations. We think the following examples will make these expressions clear.

Before giving some numerical examples, it is important to notice that it is not necessary at all to have the minimal polynomial of A. Every polynomial is good as long as it vanishes on the spectrum of A. Probably in most cases we prefer to work with $c(\lambda)$, since we know the standard method of its derivation. The reason that we introduced and worked with minimal polynomial is the ease of computation since $m(\lambda)$, as we know, has the minimum degree among all annihilated polynomials of A.

Ex. 1:

Find $f(A)$ where $A = \begin{bmatrix} 5 & -4 \\ 4 & -3 \end{bmatrix}$

$$m(\lambda) = (\lambda-1)^2 \ , \quad \lambda = 1 \text{ with multiplicity } 2$$

$\therefore$

$$f(A) = f(1)Z_{10} + f'(1)Z_{11} \quad .$$

We find Z_{10} and Z_{11} (by using particular functions $f(\lambda) = 1$ and $f(\lambda) = \lambda - 1$) from the following determinant.

$$\begin{vmatrix} f(A) & f(1) & f'(1) \\ I & 1 & 0 \\ A-I & 0 & 1 \end{vmatrix} = 0$$

$$\Rightarrow \quad f(A) - If(1) + (A-I)(-f'(1)) = 0$$

$$\Rightarrow \quad f(A) = f(1)I + f'(1)(A-I) \quad .$$

Now let us compute $f(A) = A^{200}$, using $f(\lambda) = \lambda^{200}$ we get

$$A^{200} = I + 200(A-I) = 200A - 199I \ .$$

The following computation shows that $e^{iA} = \cos A + i \sin A$, $(i = \sqrt{-1}\ )$.
To compute $f(A) = e^{-iA}$, let $f(x) = e^{-ix}$; therefore,

$$f(A) = e^{-iA} = e^{-i}I - ie^{-i}(A-I)$$

To compute $f(A) = \cos A$, let $f(x) = \cos x$; hence

$$f(A) = \cos A = \cos\,(1)I - \sin\,(1)(A-I)$$

Similarly for $f(A) = i \sin A$ we have

$$f(A) = i \sin A = i \sin(1)I + i \cos(1)(A-I)$$

$\therefore$

$$\cos A + i \sin A = I(\cos(1) + i \sin(1)) + (A-I)(i \cos(1) - \sin(1))$$

$$= Ie^{-i} + (A-I)(-ie^{-i}) = e^{-iA}$$

Q.E.D.

Ex. 2:

Find $f(A)$ where $A = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 3 & -2 \\ 4 & 1 & 0 \end{bmatrix}$

$$m(\lambda) = (\lambda-1)^2(\lambda-3), \text{ hence } \lambda_1 = 1 \text{ with multiplicity } 2, \quad \lambda_2 = 3$$

$\therefore$

$$f(A) = f(1)Z_{10} + f'(1)Z_{11} + f(2)Z_{20}$$

We find Z_{10}, Z_{11} and Z_{20} from the following determinant by using particular functions $f(\lambda) = 1$, $f(\lambda) = (\lambda-1)$, $f(\lambda) = (\lambda-1)^2$.

$$\begin{vmatrix} f(A) & f(1) & f'(1) & f(3) \\ I & 1 & 0 & 1 \\ A-I & 0 & 1 & 2 \\ (A-2I)^2 & 0 & 0 & 4 \end{vmatrix} = 0$$

$$\Rightarrow \quad 4f(A) - 4f(1)I - 4f'(1)(A-I) - (f(3)-f(1)-f'(1))(A-3I)^2 = 0$$

$$\Rightarrow f(A) = f(1)\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix} + f'(1)\begin{bmatrix}1 & -1 & 1\\3 & 2 & -2\\4 & 1 & -1\end{bmatrix}$$

$$- \frac{[f(1)+2f'(1)-f(3)]}{4}\begin{bmatrix}2 & 2 & -2\\-1 & -5 & 9\\-13 & -7 & 11\end{bmatrix}$$

$$= f(1)\tfrac{1}{4}\begin{bmatrix}2 & -2 & 2\\1 & 9 & -9\\13 & 7 & -7\end{bmatrix} + f'(1)\tfrac{1}{2}\begin{bmatrix}0 & -4 & 4\\7 & 9 & -13\\21 & 9 & -13\end{bmatrix}$$

$$+ f(3)\tfrac{1}{4}\begin{bmatrix}2 & 2 & -2\\-1 & -5 & 9\\-13 & -7 & 11\end{bmatrix} .$$

As a check set $f(\lambda) = \lambda$, i.e. $f(A) = A$; we see that our computation is correct.

The trigonometric functions of a matrix satisfy the usual trigonometric identities except when illegal matrix operations are involved such as those identities requiring negative powers of a singular matrix function. By the means of the theorem given in page 6 or formula (5) and in general by the formula (14) one can prove the following three corollaries.

Corollary 1:

The matrix $\sin \pi A$ ($\cos \pi A$) is singular $\Leftrightarrow$ A has at least one eigenvalue which is an integer (odd multiple of $\frac{1}{2}$).

Corollary 2:

The matrix $\sin \pi A$ ($\cos \pi A$) is the null matrix if A has n distinct characteristic values, all of which are integers (odd multiple of $\frac{1}{2}$).

Corollary 3:

e^A (A is an $n \times n$ matrix) is an invertible $n \times n$ matrix regardless of whether A is invertible or not.

It is useful to observe the following relations in the context of exponential function.

$$e^{A+B} = e^A \cdot e^B \quad \text{if} \quad AB = BA$$

in particular case, $B = -A$

$$e^{A+(-A)} = e^A \cdot e^{-A} = e^0 = 1$$

and

$$(e^A)^{-1} = e^{-A}.$$

We also state the following theorem without proof.

Theorem:

Every $n \times n$ invertible (complex) matrix has a logarithm, i.e., given A invertible, can find B such that $e^A = B$.

Equivalently:
V n-dimensional complex vector space

$$T: V \to V$$

then there is an

$$S: V \to V$$

such that $e^S = T$. Second form of this theorem is much better for proof.

Subspace and a basis for all defined functions of the given matrix A:

If we regard an $n \times n$ matrix as a vector in space V^{n^2} of dimension n^2, then from (14) if we prove that all the matrices Z_{ij} of the given matrix A are linearly independent, hence we can conclude that all the defined functions of the given matrix A form an m-dimensional subspace of V^{n^2} with basis Z_{ij} $(i = 1,2,\cdots,s;\ j = 1,2,\cdots,m_i-1)$.

To prove $Z_{ij}(i = 1,2,\cdots,s;\ j = 0,1,2,\cdots,m_i-1)$ are linearly independent, first we prove that $h_{ij}(\cdot)$ $(i = 1,2,\cdots,s;\ j = 1,2,\cdots,m_i-1)$ are linearly independent. For suppose that

$$\sum_{i=1}^{s} \sum_{j=0}^{m_i-1} C_{ij}\, h_{ij}(\cdot) = 0. \tag{15}$$

We show this implies that $C_{ij} \equiv 0$ for all "i" and "j". Now determine interpolation polynomial $r(\lambda)$ from the m conditions;

$$r^{(j)}(\lambda_i) = C_{ij} \qquad (i = 1,2,\cdots,s;\ \ j = 1,2,\cdots,m_i). \tag{16}$$

Then from (12) and (16)

$$r(\lambda) = \sum_{i=1}^{s} \sum_{j=0}^{m_i-1} C_{ij}\, h_{ij}(\lambda) = 0$$

but then from (16)

$$C_{ij} \equiv 0 \qquad (i = 1,2,\cdots,s; \quad j=0,1,2,\cdots,m_i-1).$$

Now it is easy to prove that Z_{ij} are linearly independent. For suppose that,

$$\sum_{i=1}^{s} \sum_{j=0}^{m_i-1} C_{ij}\, Z_{ij} = 0 \tag{17}$$

then $\Delta(A) = 0$ by definition of Z_{ij} where

$$\Delta(\lambda) = \sum_{i=1}^{s} \sum_{j=0}^{m_i-1} C_{ij}\, h_{ij}(\lambda). \tag{18}$$

But since the degree of $\Delta(\lambda)$ is less than m, the degree of the minimal polynomial $m(\lambda)$, it follows from $\Delta(A) = 0$ that $\Delta(\lambda) = 0$. Now, since the m polynomials $h_{ij}(\cdot)$ are linearly independent (19) implies that

$$C_{ij} \equiv 0 \qquad (i = 1,2,\cdots,s; \quad j=1,2,\cdots,m_i-1)$$

and this is what we had to prove.

The vector $f(A)$ has as its coordinate the m values of the function $f(\lambda)$ on the spectrum of A with respect to Z_{ij} basis.

REFERENCES

Birkhoff, G., and MacLane, S. (1965). "Survey of Modern Algebra". The Macmillan Co., New York.

Ferrar, W. L. (1951). "Finite Matrices". Oxford University Press, Amen House, London E.C.4.

Frame, J. S. (1964). "Matrix Functions and Applications". Reprinted from March-July, 1964 issues of IEEE Spectrum, Parts II and IV.

Gantmacher, F. R. (1959). "The Theory of Matrices". Chelsea Publishing Co., New York.

Hoffman, K. and Kunze, R. (1961). "Linear Algebra". Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Holmes, P. R. (1958). "Finite-Dimensional Vector Spuces". D. Van Nastran Co., Inc., Princeton, New Jersey.

Lang, S. (1966). "Linear Algebra". Addison-Wesley Publishing Co., Inc., Reading, Massachusetts.

Mitrinović, D. S., and Potts, R. B. (1965). "Elementary Matrices". P. Noordhoff, Ltd., Groningen, The Netherlands.

Narayan, Shanti (1962). "A Text Book of Matrices". S. Chand Co., Delhi.

Nering, E. D. (1963). "Linear Algebra and Matrix Theory". Wiley and Sons, Inc., New York.

Pease, M. C. (1965). "Method of Matrix Algebra". Academic Press, New York.

Pipes, L. A. (1963). "Matrix Method for Engineering". Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Schanuel, S. H. (Professor of Mathematics at Cornell University). Lecture Notes for Math. 431 and personal discussion.

Searle, S. R. (1966). "Matrix Algebra for the Biological Sciences". Wiley and Sons, Inc., New York.

Shields, P. C. (1964). "Linear Algebra". Addison-Wesley Publishing Co., Inc., Reading, Massachusetts.

Wade, T. L. (1951). "The Algebra of Vectors and Matrices". Addison-Wesley Press, Inc., Cambridge, Massachusetts.